International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:11, No:8, 2017

The Role and Importance of Genome Sequencing in Prediction of Cancer Risk

M. Sadeghi, H. Pezeshk, R. Tusserkani, A. Sharifi Zarchi, A. Malekpour, M. Foroughmand, S. Goliaei, M. Totonchi, N. Ansari–Pour

***Abstract*—**The role and relative importance of intrinsic and extrinsic factors in the development of complex diseases such as cancer still remains a controversial issue. Determining the amount of variation explained by these factors needs experimental data and statistical models. These models are nevertheless based on the occurrence and accumulation of random mutational events during stem cell division, thus rendering cancer development a stochastic outcome. We demonstrate that not only individual genome sequencing is uninformative in determining cancer risk, but also assigning a unique genome sequence to any given individual (healthy or affected) is not meaningful. Current whole-genome sequencing approaches are therefore unlikely to realize the promise of personalized medicine. In conclusion, since genome sequence differs from cell to cell and changes over time, it seems that determining the risk factor of complex diseases based on genome sequence is somewhat unrealistic, and therefore, the resulting data are likely to be inherently uninformative.

***Keywords*—**Cancer risk, extrinsic factors, genome sequencing, intrinsic factors.

I. Introduction

PERSONALIZED medicine is based on using genetic information about a person's disease to diagnose or treat his/her disease. Personalized techniques such as genome sequencing can reveal mutations in DNA that influence diseases ranging from cystic fibrosis to cancer. Personalized medicine can also be used to predict a person's risk for a particular disease allowing the physician to initiate preventative treatment before the disease presents itself in their patient. By having a detailed account of an individual's DNA sequence, their genome can then be compared to a reference genome to assess the existing genetic variations that can account for possible diseases. However, individual's health especially for complex diseases stems from genetic variation with behaviors and influences from the environment. It is believed that both intrinsic and extrinsic factors are responsible in the development of complex diseases, but the relation of each factor is not clear. The most accepted models for diseases such as cancers are based on the occurrence and accumulation of random mutational events during stem cell division. In this paper, we want to criticize these models and demonstrate that cancer risk cannot be determined using genome sequencing. In fact, a unique genome sequence assigning to any given individual (healthy or affected) is not meaningful, and hence, current whole-genome sequencing approaches are unlikely informative for personal medicine in complex diseases.

Following the divisive publication by Tomasetti and Vogelstein [1] of data indicating that the majority of variation in cancer risk is due to "bad luck", there have been a plethora of articles supporting or refuting this assertion [2]-[12]. Those who argue against the bad luck hypothesis state that this controversial claim not only concerns health organizations about a change in the perception of people towards cancer risk factors and a possible public disillusionment with the current prophylactic efforts, but also in terms of statistical analysis it is questionable, and therefore, the inference that nearly two-third of cancers are due to intrinsic (i.e. non-environmental) factors is flawed. In the most recent of these articles, Wu et al. [13] developed a statistical model based on mutation rate and number of stem cell divisions in different tissues and in contrast concluded that at least 70% of cancers are due to extrinsic factors, and intrinsic factors account for a maximum of 30% of the variation. Hence, the risk of cancer is 'less likely' to be due to bad luck.

Analyses in Wu et al. [13] were based on exactly the same lifetime risk of cancer data as in Tomasetti and Vogelstein [1], obtained from the Surveillance, Epidemiology and End Results (SEER) Program database [14]. Both studies assumed that genetic mutation is the underlying cause of cancer, and thus the larger the number of tissue-specific stem cells and cell divisions, the higher the rate of mutation accumulation and hence, a higher risk of cancer. Interestingly, both studies agree that intrinsic factors, independent of the extrinsic factors, play a role in cancer etiology. The schism between the two is merely based on the magnitude of cancer risk estimated to pertain to these factors. One can thus argue that cancer, notwithstanding the level of interplay between intrinsic and

M. Sadeghi is with the National Institute of Genetic Engineering and Biotechnology, Tehran, Iran and School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM) (corresponding author, phone: 21-44787373, e-mail: msadeghi@ipm.irauthor

H. Pezashk and A. Malekpour are with the School of Mathematics, Statistics and Computer Science, College of Science, University of Tehran, Tehran, Iran (e-mail: pezeshk@khayam.ut.ac.ir, a.malekpour@ut.ac.ir).

R. Tusserkani is with the School of Computer Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran (e-mail: ruztus@yahoo.com).

A. Sharifi Zarchi is with the Department of Stem Cells and Developmental Biology, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran (e-mail: asharifiz@gmail.com).

M. Foroughmand is with the Department of Mathematical Science, Sharif University of Technology, Tehran, Iran (e-mail: foroughmand@gmail.com).

S. Goliaei and N. Ansari–Pour are with the Faculty of New Sciences & Technology, University of Tehran, Tehran, Iran (e-mail: sgoliaei@ut.ac.ir, n.a.pour@gmail.com).

M.Totonchi is with the Department of Genetics, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran (e-mail: totonchimehdi@gmail.com).

extrinsic factors, is a stochastic phenomenon and therefore inherently has a bad luck component.

II. Extrinsic and Intrinsic Risk

A. Bad Luck and Risk

It seems to us that the notion of bad luck in cancer etiology is not fully understood and should be clarified. When a phenomenon is said to be stochastic, it means that we cannot predict its outcome in advance. In the other words, there is uncertainty in predicting such an outcome. This concept is independent of the degree of chance and key to understanding the notion of bad luck. To elaborate, if we lose in a lottery where there is a 70% chance of winning, we are said to have experienced bad luck. Indeed, losing in a lottery where the chance of winning is only 30% is also bad luck since winning in either situation is considered to be 'good luck'. Therefore, losing (e.g. developing cancer) is thought to occur randomly irrespective of the degree of chance of losing (30% or 70%). To put this in context, it is palpable that environmental factors can increase or decrease the chance of developing cancer but this does not contradict the concept of bad luck. For instance, with the absence of extrinsic factors (which can unequivocally increase the mutation rate), the occurrence of cancer is indeed due to intrinsic factors only and purely stochastic due to the random nature of genetic mutation. However, in the presence of extrinsic factors, the chance of cancer occurrence increases when compared with individuals who are not subject to these factors. Nevertheless, importantly, the random nature of its occurrence remains well present. This can be further exemplified if we were to look at the risk of cancer development in a cohort of smokers since everyone would be equally likely to get cancer due to bad luck.

Moreover, although the two studies assign different weights to intrinsic and extrinsic factors, their inferences are based on the same hypothesis that the risk of cancer is directly proportional to the number of stem cell divisions in a given tissue. In this model, it is assumed that during each cell division, mutation occurs at a specific rate and tissues with larger number of stem cells, and hence divisions, accumulate mutations at a higher rate. This higher mutation accumulation rate increases the chance of mutations occurring in driver genes. The effect of environmental factors in increasing the mutation rate and in turn increasing cancer risk in a particular tissue or cell type does not contradict the effect of number of stem cell divisions and the random nature of cancer occurrence but only biases the degree of chance. In the other words, it is in fact equivalent to a longer lifespan or more division cycles of stem cells. Consequently, if we were to only consider cancer patients who are exposed to environmental risk factors (such as UV radiation and smoking), we would expect the number of stem cell divisions and cancer incidence to be highly correlated. Therefore, intrinsic factors are at play in all cases, and the concept of bad luck remains core to cancer occurrence.

B. Age and Risk

We believe that analyses in both studies are not quite representative and adequate to estimate accurately the weight of intrinsic and extrinsic factors in cancer occurrence. For instance, in their model, it is assumed that the number of stem cells is initially increased exponentially through m symmetric divisions to 2m cells, and then, with a constant rate, followed by asymmetric divisions in which one of the two daughter cells remains a stem cell. According to cell-specific division rates and the time remaining after achieving 2m stem cells, n asymmetric divisions occur. In these analyses, n is estimated according to the annual cell division rate and an arbitrarily assumed average lifespan of 80 years. The total number of stem cell divisions in each tissue was then estimated by multiplying n+m divisions by the number of stem cells in that tissue.

If the proposed model is valid, we expect that lifetime cancer risk would also be directly linearly proportional to age. To test this prediction, we obtained data on the rate of cancer incidence in different age groups from the SEER database (see Table I). As depicted in Fig. 1, as the age increases, there is an increase in incidence of most cancer types. In some types of cancer, this increase is up to 75 years, but the cancer incidence in breast, female genital system, oral cavity, pharynx and skin increases up to 65 years and is then reduced. In breast (in situ) and endocrine system, the increase is up to 55 years. Although the reason for this observation is yet to be established [15], [16], the difference in cancer incidence in the older ages may be explained with decreased number of stem cells at a specific age (after a period of constant renewal) with this age being tissue-specific. Therefore, the linear regression model of lifetime risk of cancer on number of divisions up to 80 years (i.e. total estimated number of cell divisions) is unlikely to be a realistic model. Moreover, we observed an interesting contrast in cancers of bones and joints, brain and nervous system, eye and orbit, and leukemia, which show high incidence up to 20 and above 40 years of age but occurs at a much reduced rate between 20 and 40 years. These data suggest that employing a simple linear regression model based on the number of stem cells and their asymmetric divisions (and mutation rate) to estimate lifetime cancer risk is far from realistic. It importantly also indicates that we cannot expect a similar lifetime risk of cancer for different tissues based on similar estimated number of divisions. Indeed, this forms the central criticism of Wu et al. [13] to the hypothesis of Tomassetti & Vogelstein [1] by focusing on tissues with unequal lifetime risk of cancer but with similar cell division number and justifying the differential risk based only on extrinsic factors. However, we provide another hypothesis based on age-related cancer incidence. After birth, number of stem cells and their divisions are expected to rise and continue with higher rates in certain tissues during childhood. This results in an increased chance of cancer incidence due to mutation accumulation consistent with the high incidence of cancer in such tissues below the age of 20. Furthermore, in addition to different cell division rates at different ages, different levels of molecular activities (e.g. metabolism, inter-

cellular communication and the local environment) within the same tissue at different stages of the lifespan may also affect cancer incidence. Since the relationship of cancer incidence rate and its relation with different ages may be unique to the US population, we also analyzed data from the United Kingdom [17], and the rate of different cancers in different age groups showed a similar pattern to that observed in the US. Ignoring unequal rates of cancer incidence across different age groups, as done in previous analyses, makes the linear regression model less reliable. We thus demonstrate that the role of extrinsic and intrinsic factors is unlikely to be precisely and concretely determined solely based on the SEER data and those alike.

TABLE I
PERCENT OF DIFFERENT CANCERS IN DISTINCT AGE GROUPS IN THE UNITED STATES

Site	<20	20-34	35-44	45-54	55-64	65-74	75-84	85+	Total Cases
All Sites	1	2.6	5.3	14.2	23.8	25.1	20.1	7.8	1968702
Oral Cavity & Pharynx	0.6	2.1	5.9	20.2	29.1	21.4	14.7	6	47286
Digestive System	0.2	1	3.6	13.4	22.2	24.4	23.8	11.3	361202
Respiratory System	0.1	0.4	1.5	9.3	21.8	31.1	27.3	8.5	274554
Bones & joints	27.4	15.3	9.7	13.1	11.9	9.5	9.3	3.6	3888
Soft tissue (including heart)	9	9.4	10	14.8	17.2	16.2	16.2	7.2	13963
Skin (excl. basal & squamous)	0.6	6.4	9.8	17.2	21.2	19.5	17.6	7.7	98066
Breast (Female)	0	1.8	9.6	22.2	25.2	20.7	14.8	5.7	285391
Breast (Female -in situ)	0	0.7	10.7	28.8	26.8	19.7	11.2	2.2	72153
Female Genital System	0.4	4.1	9.2	19.2	27.3	20.1	13.7	5.9	112905
Male Genital System	0.2	1.8	1.6	9.8	31.2	34.5	17.1	3.9	309497
Urinary System	0.6	1	3.4	11.1	21.6	26.5	25.5	10.4	154986
Eye & Orbit	13	3.2	6.3	15	20.2	20.2	15.8	6.2	3429
Brain & Nervous System:	13.1	8.8	8.8	14.9	19.5	16.7	13.4	4.8	27410
Endocrine System	2.9	14.8	19.3	23.8	19.4	12.2	6	1.5	55611
Lymphoma:	3.1	7.2	7.4	13.4	19.2	21	20.5	8.3	94880
Myeloma	0	0.5	3.2	11.4	22.8	27.4	25.2	9.6	25028
Leukemia:	10.3	4.7	5.2	10.4	16.7	20	21.7	11.1	53448
Kaposi Sarcoma	0.2	19.1	29	21.1	9.2	7.5	7.8	6.1	2283
Mesothelioma	0.1	0.6	1.9	6.4	16.1	27	34.1	13.7	4184
Ill-defined & unspecified	0.4	0.9	2.4	9.6	18.1	22.2	27.5	18.8	38395

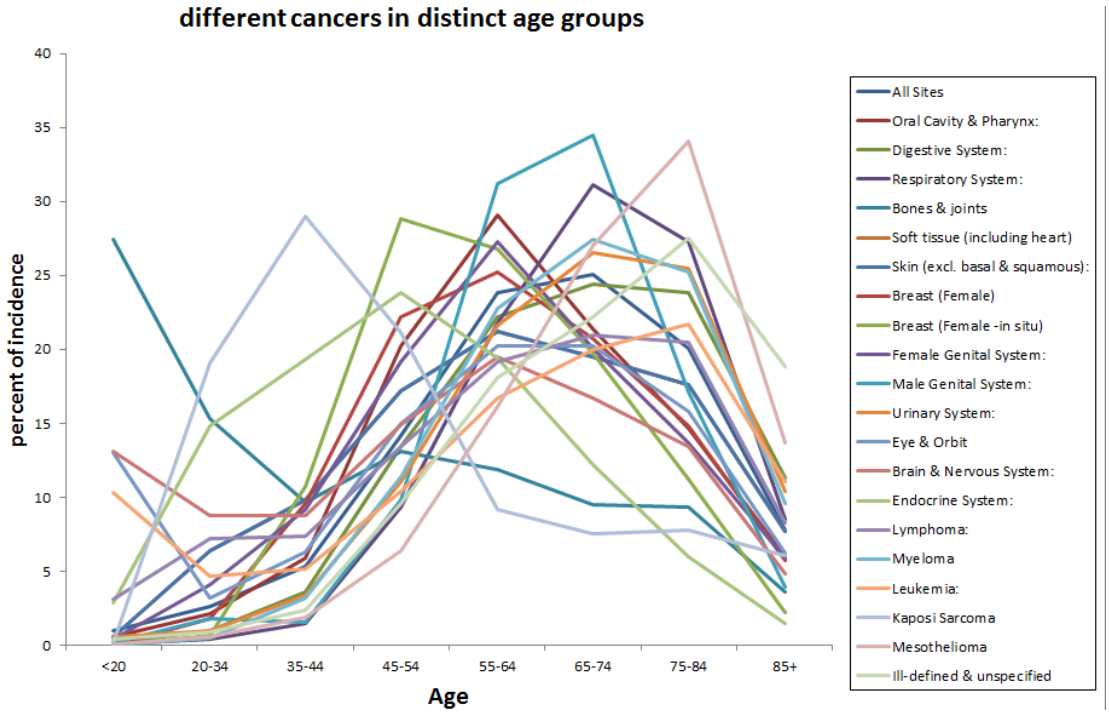

Fig. 1 Age dependency of cancer incidence in the United States. As shown, in the most cases, cancer incidence reduces after an increase up to 65 years old but this pattern is not the same in all cancers

The analogy made recently by Tomassetti and Vogelstein between developing cancer and occurrence of a car accident [18] serves well to illustrate this further. They compared the number of stem cell divisions to the length of the trip, which increases the probability of an accident. On the other hand, environmental and inherited factors are compared to road and mechanical conditions respectively, which increase the probability of an accident. However, the fact ignored in this analogy (and other cell division based models) is that stem cells specific to each tissue can be compared to cars of different motor vehicle manufacturers displaying distinct qualities and therefore different probabilities of getting into an accident. Hence, the assumption that an equal trip length on the same road results in an equal accident probability for all different cars is not realistic based on our data-driven analysis.

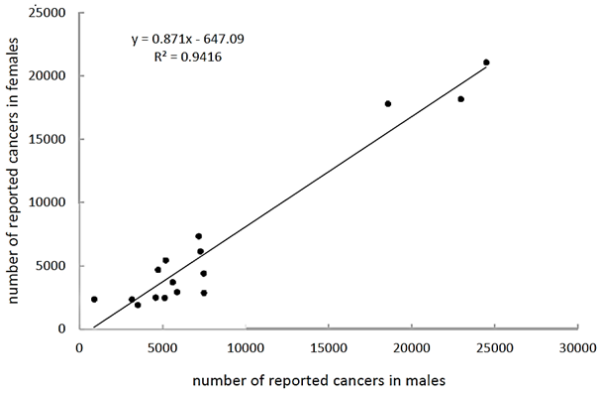

Fig. 2 Incidence of common cancers incidence is highly correlated between males and females in the United Kingdom and generally higher in males (Number of New Cases in 2013)

By considering the number of divisions of both stem cells and total cells in different tissues, Wu et al. have claimed that 70-90% of cancers are due to extrinsic factors [13]. An instance of environmental factors is the role of smoking in lung cancer. In the SEER database, in addition to the number of patients for each cancer type, the incidence has also been reported based on sex in two cancer categories of 'Lung & Bronchus' (135,823 males and 120,167 females) and 'Colon and Rectum' (97,800 males and 92,366 females). Although the slightly higher incidence of both cancer types in males may be due to chance, we compared the incidence of different cancers between males and females to see if this small difference is due to the larger size of the respective tissues in men. Fig. 2 shows that incidence in males and females is highly correlated across different cancers and generally higher (15%) in males than females, suggesting that higher number of cells may lead to an increased rate of cancer incidence. Interestingly, the pattern of age-dependent cancer incidence was similar in males and females with incidence for the 'Lung and Bronchus' category being almost identical (Fig. 3).

Consequently, if extrinsic factors (environmental factors), as predicted by Wu et al. [13], are the predominant force in increasing incidence of this cancer type, we expect to observe a similar ratio of smoking males and females. However, data on smoking patterns in the US population show that there is a higher smoking prevalence among males than females [19]. The similar incidence rate in both sexes is thus not explained predominantly by an extrinsic factor like smoking. This suggests that, to demonstrate that extrinsic factors are mainly responsible for lung cancer incidence, certain extrinsic factors are ought to be identified that both sexes have equal exposure to it. This claim also holds for the almost identical incidence of colon cancer. Otherwise, the alternative is to infer that SEER data are not suitable for estimating the contribution of extrinsic and intrinsic factors towards cancer etiology. Although data of other countries may lead to different conclusions, data on the incidence of different cancers categorised by sex in the United Kingdom were similar to those in the SEER database (Table II). Nonetheless, analysis of sex-based incidence data of other cancer types may be a powerful tool for identifying extrinsic risk factors if a significant difference in cancer incidence between males and females correlates with exposure to a risk factor. Moreover, this can be extended to comparing populations with different cultural practices. For instance, incidence data of melanoma may be compared between countries where sun tanning (exposure to UV) is practiced to a different degree.

TABLE II
THE 15 MOST COMMON CANCERS IN MALES AND FEMALES IN UK* (NUMBER OF NEW CASES IN 2013)

Cancer Site	Male	Female	Persons
Lung	24481	21044	45525
Bowel	22957	18155	41112
Malignant Melanoma	7152	7357	14509
Non-Hodgkin Lymphoma	7259	6154	13413
Kidney	7455	4418	11873
Brain, Other CNS	5164	5460	10624
Bladder	7465	2876	10341
Pancreas	4716	4692	9408
Leukaemia	5585	3716	9301
Oesophagus	5852	2932	8784
Oral	5103	2488	7591
Stomach	4564	2503	7067
Myeloma	3142	2355	5497
Liver	3491	1922	5413
Thyroid	880	2361	3241

*cruk.org/cancerstats

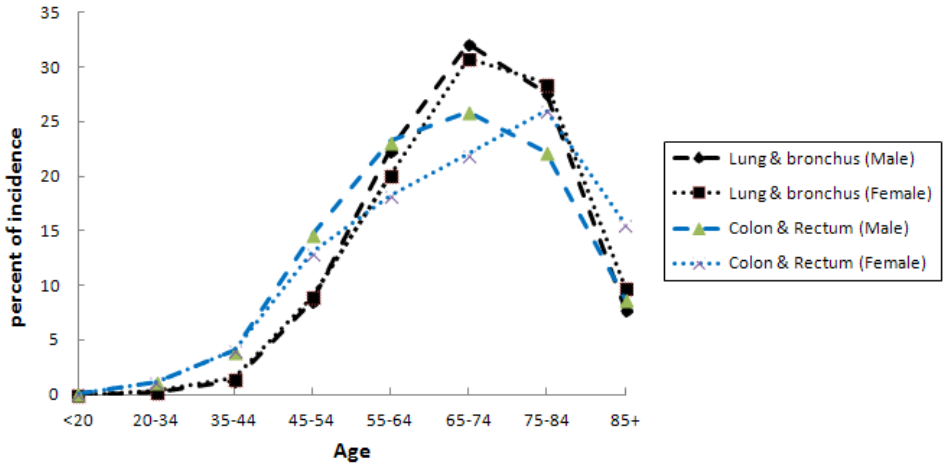

Fig. 3 The pattern of age and sex dependency for colon and lung cancer incidence based on SEER data. The pattern is similar in both males and females

III. Is Whole-Genome Sequencing Still Relevant?

Perhaps, the most important yet concealed issue in the studies mentioned above is that the germline genome of an individual is unlikely to be a major indication for cancer incidence, and access to the genome sequence of an individual does not provide tangible information about risk of cancer or the probability of occurrence in that individual. The first argument for this claim is that cancer is a mutation-driven stochastic process. Mutation accumulation is either due to intrinsic factors and directly proportional to the number of stem cell divisions or due to extrinsic factors and increased due to environmental risk factors. Consequently, sequencing the genome of an individual (either based on whole-blood or biopsy of any specific tissue) at any time during its lifespan is inadequate to predict cancer occurrence since it cannot foresee mutations occurring in the immediate future in driver genes which may lead to cancer development. Sequencing genomes of individuals for the purposes of cancer risk estimation is therefore likely to have no clinical relevance.

The second line of evidence comes from data of cancer incidence in immigrated people. It has been observed that when people immigrate from low cancer incidence countries to those with high incidence, they acquire a higher risk of cancer development in a short period of time [20], [21]. We agree with Wu et al. [13] that this may be due to the new diet and other external factors of the new environment. Accordingly, genome sequencing of these people is very unlikely to provide accurate information about their risk of cancer incidence.

The claim is further confirmed by observations made on monozygotic twins. A plethora of studies have shown that the risk pattern of developing complex diseases shows variable concordance between monozygotic twins albeit having the exactly same germline genomic sequence [22]-[25]. With the view of appraising the usefulness of personal genome sequencing, Roberts et al. [22] modeled the risk of 24 common complex diseases in a large number of monozygotic twins, and observed that, for 80% of the diseases, the relative risk for those tested negative based on genome sequencing did not differ significantly from that of the general population. They thus concluded that whole genome sequencing is not informative in predicting complex disease incidence, especially in those with a low heritable component such as many types of cancer.

Assuming that we accept the mutation-driven model of cancer, regardless of the weight of intrinsic and extrinsic factors, in which mutational events occur in each stem cell division and accumulate with increased number of divisions (and in turn increase the chance of driver mutations), personal genome sequencing seems to be inherently uninformative. For example, in the duodenum with an average of 4,000,000 stem cells, if we assume 21 symmetric divisions (m) to achieve this number of early stem cells and then 24 asymmetric divisions in each year [26], [27], each stem cell will undergo 1221 divisions during the lifespan of a 50-year-old individual (n=1200). Assuming that only point mutations occur (at a mutation rate of 10-8), at this age each diploid stem cell (and subsequent daughter cells) of the duodenum of this individual will have a minimum of 145,000 nucleotide changes. In the other words, there would be a maximum of 290,000 different nucleotide sequences in any pair of duodenum stem cells separated very early in life. These numbers are likely to be underestimated, since the analysis was based only on point mutations (single nucleotide variants), while larger and more complex variants may well occur. The main inference from this model is that assigning one specific genomic sequence to an individual is impossible when we talk about personal genomes. This is because depending on the cell(s) sampled and the time of sampling, there would be unique genome sequences for each cell in each time frame, all of which may have tens or hundreds of thousands of different nucleotide sequences.

In reality, what is referred to as the genome of an individual is in fact the consensus sequence of all cells within the sampled tissue and numerous inter-cellular sequence differences are thus masked by the common allelic variant(s).

Although it is true that the consensus sequence will include inherited and germline mutations of which some may increase the risk of cancer development (such as those in BRCA1/2 and TP53), it will not contain most of the variation accumulated in cells of a tissue over the course of time. More importantly, the consensus sequence of different tissues, especially those separated early in the embryonic stages, are likely to be different at many nucleotide positions due to the accumulation of many tissue-specific mutations. In the other words, each tissue is likely to have its own consensus sequence. Interestingly, corroborative evidence for this notion was shown in a recent exome analysis of 16 tissues of a single individual where significant intertissue variation was observed [28]. This indicates that assigning a unique sequence to an individual is even more challenging when analysing the whole genome. Yet, many sequencing projects are still continuously funded to compare the genomes of cancer patients with control individuals to identify driver mutations in cancer. This approach seems to be of limited value for identifying such mutations since the sequences compared are essentially consensus sequences of patient and control individuals and at best may identify rare inherited and germline mutations, which, in carriers, are likely to further elevate cancer risk compared with the baseline risk pertaining to somatic mutation accumulation common to all individuals. Under such circumstances, it is thus not farfetched to deem current genome sequencing approaches irrelevant for assessing the risk of developing cancer, which is a stochastic process.

We do nevertheless acknowledge that genome sequencing-based cancer projects that address this mosaicism have the potential to overcome this issue (see [29] as an example). We hope that this study provides a platform for an effective debate on the clinical relevance of genome sequencing and that the limitations of this approach in predicting cancer incidence are recognized by researchers, large genome sequencing-based consortia and major healthcare funding agencies.

IV. Conclusion

In conclusion, since the genome sequence of an individual differs from cell to cell and changes over time, it seems that the current expectation from personal genomics in determining the risk factors of complex diseases is somewhat unrealistic given that most large genome projects to date lack the resolution to identify all possible driver mutations in individuals and therefore the resulting data are likely to be inherently uninformative.

References

[1] C. Tomasetti, B. Vogelstein, "Variation in cancer risk among tissues can be explained by the number of stem cell divisions," *Science*, vol. 347, pp. 78–81, 2015.
[2] N. A. Ashford, et al., "Cancer risk: role of environment," *Science*, vol. 347, p. 727, 2015.
[3] C. Wild, P. Brennan, M. Plummer, F. Bray, K. Straif, J. Zavadil, "Cancer risk: role of chance overstated,'' *Science*, vol. 347, p.728, 2015.
[4] J. D. Potter, R. L. Prentice, "Cancer risk: tumors excluded," *Science*, vol. 347, p. 727, 2015.
[5] C. Gotay, T. Dummer, J. Spinelli, "Cancer risk: prevention is crucial," *Science*, vol. 347, p. 728, 2015.
[6] M. Song, E. L. Giovannucci, "Cancer risk: many factors contribute," *Science*, vol. 347, pp. 728–729, 2015.
[7] M. O'Callaghan, "Cancer risk: accuracy of literature," *Science*, vol. 347, p. 729, 2015.
[8] J. T. Leek, R. D. Peng, "What is the question? Mistaking the type of question being considered is the most common error in data analysis," *Science*, vol. 347, pp. 1314-1315, 2015.
[9] D. Wodarz, A. G. Zauber, "Cancer: Risk factors and random chances." *Nature*, vol. 517, pp. 563–564, 2015.
[10] C. R. Weinberg, D. Zaykin, "Is Bad Luck the Main Cause of Cancer?" *JNCI J Natl Cancer Inst*, vol. 107, djv125, 2015.
[11] A. I. Rozhok, G. M. Wahl, J. A. DeGregori, "Critical Examination of the "Bad Luck" Explanation of Cancer Risk," *Cancer Prev Res,* vol. 8, pp. 762–764, 2015.
[12] C. C. Harris, "Cause and Prevention of Human Cancer," *Carcinogenesis,* vol. 36, S1, 2015.
[13] S. Wu, S. Powers, W. Zhu, Y. A. Hannun, "Substantial contribution of extrinsic risk factors to cancer development," *Nature*, vol. 529, pp. 43-47, 2015.
[14] National Cancer Institute, Surveillance, Epidemiology, and End Results Program; Accessed on 01/11/2016, Sub (1973-2014) www.seer.cancer.gov.
[15] J. K. Pedersen, *et al*. "Cancer and aging: Epidemiology and methodological challenges," *Acta Oncologica,* vol. 55, pp. 7-12, 2016.
[16] P. D. Adams, H. Jasper, K. L. Rudolph, "Aging-Induced Stem Cell Mutations as Drivers for Disease and Cancer," *Cell Stem Cell,* vol. 16, pp. 601-612, 2015.
[17] A. S. Ahmad, N. Ormiston-Smith, P. D. Sasieni, "Trends in the lifetime risk of developing cancer in Great Britain: Comparison of risk for those born in 1930 to 1960," *Br J Cancer*, vol. 3, pp. 943-647, 2015.
[18] J. E. Lucas, G. Sapiro, "Cancer: What's luck got to do with it?" *Significance*, vol. 12, pp. 40–42, 2015.
[19] Centers for Disease Control and Prevention. Current Cigarette Smoking Among Adults—United States, 2005–2014. Morbidity and Mortality Weekly Report, vol. 64, pp. 1233–1240, 2015.
[20] J. Gray, & et al. "State of the evidence: the connection between breast cancer and the environment," *Int. J. Occup. Environ. Health*, vol. 15, pp. 43–78, 2009.
[21] H. Shimizu, R. K. Ross, L. Bernstein, R. Yatan, B. E. Henderson, "Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County," *Br. J. Cancer*, vol. 63, pp. 963–966, 1991
[22] N. J. Roberts, J. T. Vogelstein, G. Parmigiani, K. W. Kinzler, B. Vogelstein, V. E. Velculescu, "The predictive capacity of personal genome sequencing," *Sci Transl Med*, vol. 4, pp. 133-158, 2012.
[23] C. S. Haas, C. J. Creighton, X. Pi, I. Maine, A. E. Koch, G. K. Haines, S. Ling, A. M. Chinnaiyan, J. Holoshitz, " Identification of genes modulated in rheumatoid arthritis using complementary DNA microarray analysis of lymphoblastoid B cell lines from disease-discordant monozygotic twins," *Arthritis Rheum*, vol. 54, pp. 2047-2060, 2006.
[24] K. A. Metcalfe, G. A. Hitman,R. E. Rowe, M. Hawa, X. Huang, T. Stewart, R. D. Leslie, "Concordance for type 1 diabetes in identical twins is affected by insulin genotype," *Diabetes Care*, vol. 24, pp. 838-842, 2001.
[25] W. Czyz, J. M. Morahan, G. C. Ebers, S. V. Ramagopalan, "Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences," *BMC Medicine*, vol. 10, p. 93, 2012.
[26] J. Y. Kim, K. D. Siegmund, S. Tavaré, D. Shibata, "Age-related human small intestine methylation: Evidence for stem cell niches," *BMC Med*, vol. 3, p. 10, 2005.
[27] S. Kozar, E. Morrissey, A. M. Nicholson, M. van der Heijden, H. I. Zecchini, R. Kemp, S. Tavaré, L. Vermeulen, D. J. Winton, "Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas," *Cell Stem Cell*, vol. 13, pp. 626–633, 2013.
[28] A. Gómez-Ramos, R. Sanchez-Sanchez, A. Muhaisen, A. Rábano, E. Soriano, J. Avila, "Similarities and Differences between Exome Sequences Found in a Variety of Tissues from the Same Individual," *PloS One*, vol. 9, e101412, 2014.
[29] E. Ruark, et al. "Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer," *Nature*, vol. 493, pp. 406-410, 2013.